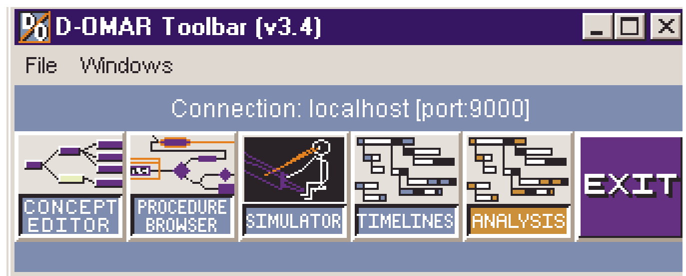
BBN Report No. 8328

Modeling Human Error in D-OMAR

Stephen Deutsch
Richard Pew



December, 2001

Prepared by: BBN Technologies
10 Moulton St.
Cambridge, MA 02138

Prepared for: Aviation Safety Program
NASA Ames Research Center
Moffett Field, CA 94035

Table of Contents

1. Introduction

Human performance modeling is a technology that has the potential for addressing flight deck and ATC workplace design and procedure issues early in the development process before prototypes or full-scale simulations are available. However, the technology is not yet sufficiently mature that one can apply the technology routinely. At this stage, it is possible to undertake exploratory investigations that can highlight the potential for improving the cost-effectiveness of the design process and at the same time strengthen the architectures and tools that are available for the modeling process. Under the aegis of the Aviation Safety Program, NASA ARC has initiated a program element concerned with modeling human error in order to advance the state-of-the-art and to demonstrate potential payoffs in the commercial aircraft design world. This element is a multi-year, multi-contractor effort that is just completing its first year. For the first year effort, NASA chose to focus on ground operations. BBN's effort, for which this is the First Year Final Report, has focused on approach and landing, and taxi operations, with an emphasis on the interactions among the air traffic controllers, the captain and first officer and the aircraft controls and displays.

NASA has provided BBN with the detailed results of the Taxiway Navigation and Situation Awareness (T-NASA) experiments (NASA, 2001a through 2001i). The purpose of the two experiments was to evaluate the usefulness of advanced cockpit displays specifically designed to support low-visibility taxi operations. The experiments included a baseline condition using current instrumentation, and several configurations of the T-NASA head-up and head-down displays. The T-NASA displays led to nearly error-free trials—the focus of the BBN human error modeling effort was on the baseline conditions. In the first study, 16 two-pilot commercial crews completed six land-and-taxi-to-gate trials based on current operations using the simulated terrain of Chicago O'Hare Airport. In the second study, 18 commercial two-pilot crews completed three land-and-taxi-to-gate trials based on current operation equipment. All trials were conducted under low visibility or nighttime operating conditions. The crews consisted of a sample of airline captains and first officers, both current in the same aircraft type and from the same airline, with varied levels of experience and familiarity with O'Hare. In evaluating these studies, Hooey and Foyle (2001) defined navigation errors as taxiing on a portion of the airport surface on which the aircraft had not been cleared and deviating from their cleared centerline by at least 50 feet. Their analysis revealed 26 navigation errors in 150 current-operation trials—errors were committed on 17.3% of the trials. Our naïve view was that while failures may occur with modest frequency, most are readily caught and very, very few lead to incidents—certainly not at the rates reported.

The modeling effort was accomplished using the Distributed Operator Model Architecture[1] (D-OMAR), a BBN human performance modeling environment, to represent the behaviors of the aircrews and air traffic controllers. D-OMAR was also used to model the aircraft and their flight decks, the ATC workplaces, and the essential features of the Chicago O'Hare Airport. Our goal was to produce a model that could

[1] The D-OMAR software with user manual is available as OpenSource at http://omar.bbn.com.

appropriately represent successful taxi performance and with modest adjustments represent a subset of the classes of error that were observed in the T-NASA experiments.

After an introduction to the aviation context in Section 2, this report presents, in Section 3, the BBN approach to human modeling represented in D-OMAR. Because we take the view that errors are the result of small variations from nominal performance, in Section 4 we next discuss our approach to human error in general and, more specifically, the ways that errors might arise in the context of otherwise successful human performance. This section describes many more ways that errors could arise than we have actually explored thus far in the modeling experiments. Then in Section 5 we describe specifically the error modeling experiments on taxi operations we have completed in this first year of the project.

2. The Approach-and-Landing and Taxi Context

D-OMAR provided the simulation framework for the implementation of the T-NASA baseline approach-and-landing and taxi scenarios. Human performance models were developed for each aircraft's captain and first officer and for the approach and ground controllers. A single aircraft satisfied the requirements for a majority of the simulation runs. In one of the error sequences, a second aircraft was used to provide taxiway traffic to increase the workload for the primary aircraft's aircrew.

The D-OMAR simulator is a very efficient event-based simulator. When running in fast-time, as was done most often, the scenarios executed in about one tenth of real time. Simulation objects were defined in the Simple Frame Language (SFL), a classical frame language derived from KL-ONE (Brachman & Schmolze, 1985). Object behaviors, including the human performance models, were defined using the Simulation Core (SCORE) language, a procedural language. SFL and SCORE are languages developed in Lisp. The D-OMAR simulation engine that executes agent behaviors is also written in Lisp, while the user interface tools used in building the models, monitoring runtime activities, and for post-run analysis are implemented in Java. Model development was done in a Window NT environment. D-OMAR will run most Windows, Unix, and Linux machines where Franz Lisp and Java are available. Franz Lisp has recently become available on the Macintosh, but no attempt has yet been made to run D-OMAR there. D-OMAR and the human error models developed for this research effort can be made available to run in any of these operating environments.

The Nominal Task Sequence (NASA, 2001a) provided a sequential listing of the nominal tasks performed by aircrews during the T-NASA-2 scenarios. The sequence was derived from a sample of eight trial runs and were selected to fairly reflect the requirements of the baseline simulation conditions. As such, the task sequence established the requirements for the basic elements of the simulation: the captain, first officer, and approach and ground controller capabilities and procedures; aircraft model capabilities; flight deck systems; ATC workplace systems; and the Chicago O'Hare Airport component models.

As described in the Nominal Task Sequence, the scenarios began about twelve miles out in the approach sequence. The aircrew was engaged in intra-crew communication, and the aircrew and the approach controllers were engaged in party-line communication related both to the approach and landing sequence and the subsequent taxi procedures. The crew followed a basic approach sequence. The captain set the approach speed to 180 knots, engaged speed mode, and made the appropriate call-outs attended to the first officer. The approach controller then provided the landing clearance and preferred exit. These were discussed by the captain and first officer and subsequently acknowledged by the first officer who then wrote down the preferred exit. The crew monitored the approach profile, and verified localizer and glideslope capture. The aircrew made and attended to the related call-outs.

Next, the crew configured the aircraft for landing. The captain called for “gear-down,” the first officer engaged the landing gear and made the call-out. The captain adjusted the speed to 146 knots, made the call-out, and then called for “flaps 25.” The first officer set the flaps to 25 and made the call out. The captain then called for the final descent checklist. The first officer verified that the spoilers were armed, the landing gear was down, and the flaps were set to 25 degrees, made the call-outs and confirmed the completion of the checklist.

At this point, the aircrew was monitoring the final descent profile. The first officer monitored the aircraft’s altitude and made call-outs for 1000 feet, 500 feet, and “minimums.” The captain announced “landing,” the first officer called out 100 feet, and the aircraft touched down. The captain and first officer “sensed” weight-on-wheels, the captain applied reverse thrust and disarmed the autobrakes. The first officer verified the thrust levers were closed and the speedbrakes were up. The first officer then monitored the aircraft’s ground speed and made call-outs at 100, 90, 80, 70, and 60 knots. At this point, the landing was complete and transition to taxi operations was about to take place. The aircraft model provided each of the controls required by approach-and-landing sequence as outlined here. The aircrew executed each of the required steps and monitored the approach as indicated, however the aircraft model was simply following an open loop landing trajectory.

It was important to model the approach in a manner that closely followed actual operations because it was early in the approach sequence that the first information on the subsequent taxi operations was obtained and processed by the aircrew. Even though the aircrew was then totally absorbed in safely executing the landing procedure, this early taxi information was processed to a level that it might well influence upcoming taxi operations and it was this influence that it was important to capture.

As the aircraft slowed to taxi speed, the aircrew prepared to take the preferred exit from the runway with the captain in closed-loop active control of taxi operations. The first officer reviewed his/her notes and the airport diagram and notified the captain of the aircraft’s location with respect to the preferred exit. The captain disengaged the autopilot and monitored the runway signage for the runway exit. At the aircraft approached the exit, the first officer notified the controller that the aircraft was clearing the runway, the aircrew switched their radio frequencies to that of the ground controller and the first officer notified him/her that the aircraft had cleared the runway. The ground controller responded with the taxi clearance. The captain monitored the communication while first officer took notes on the taxi sequence.

The aircrew then followed a basic procedural pattern in executing the taxi operations. As each turn was completed, the first officer reviewed his/her notes for next turn in the taxi sequence, checked the airport diagram for turn location and taxiway geometry, and notified the captain of the upcoming turn providing additional details as the required by the geometry of the taxiway layoout. As the captain controlled the progress of the aircraft, the captain and first officer (when not head-down) scanned the out-the-window view for traffic. The captain monitored and called out the airport signage, tracked the current centerline for the next turn, and announced the turn as he/she executed it. The pattern then repeated until the final taxiway before the approach to the concourse gate at which point the scenario terminated. Section 4.3 provides additional detail on the taxi procedures and in particular focuses on the modeling of error as developed in this research effort.

Figure 1 provides a screen shot showing the progress of NASA186 as it nears the completion its taxi operations. The orange line represents Chicago O'Hare runway 9R/27L and the blue lines represent a subset of the taxiways leading to concourses H, K, and L. The panels on the right provide details on aircrew and controller communications. The top panel shows the radio dialog between NASA186 and the approach controller, the middle panel, details subsequent radio communication between NASA186 and the ground controller, and the bottom panel provides details of the intra-crew conversation on the NASA186 flight deck. Runway and signage data for the model was derived from NASA (2001h) provided information.

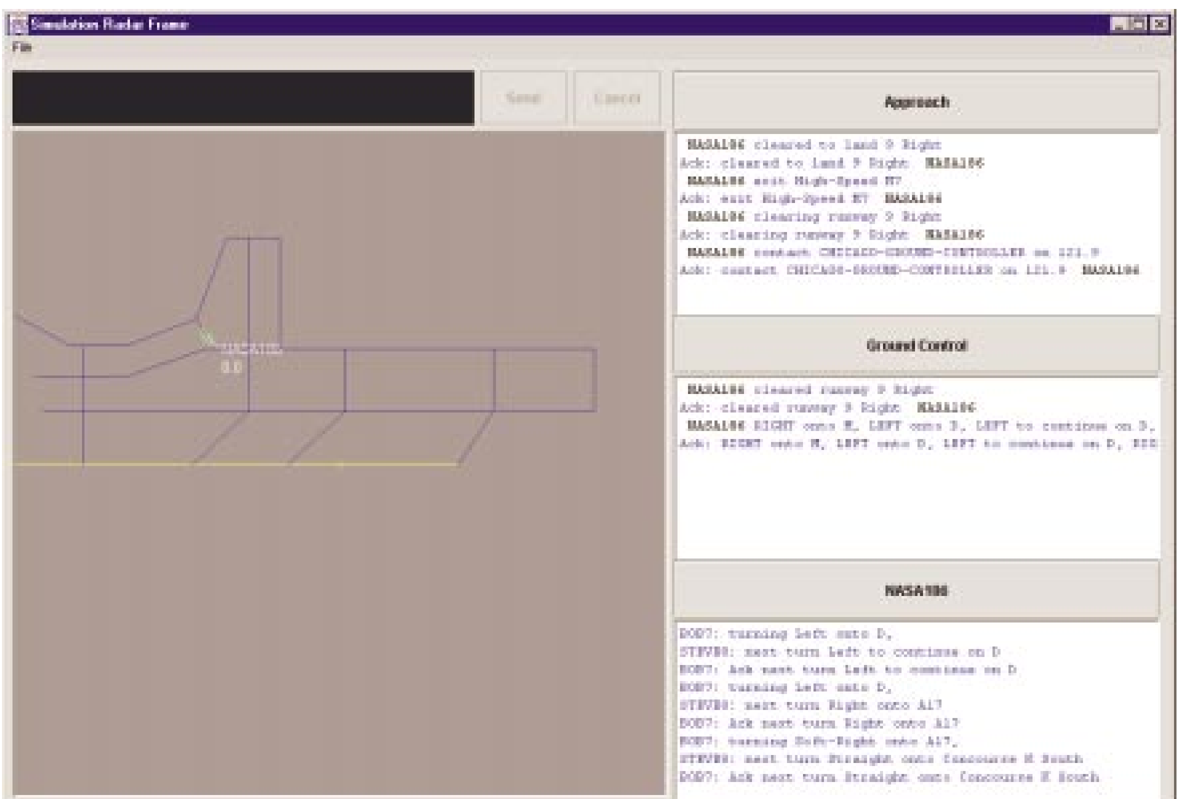


Figure 1. D-OMAR Human Error Modeling Scenario

3. Human Modeling in D-OMAR

From their inception, the human models developed in D-OMAR (Deutsch, 1998; Deutsch & Adams, 1995) have been grounded in research in cognitive neuroscience, cognitive science, experimental psychology, and recent cross-disciplinary work in the theory of consciousness. Neumann's (1987) functional view of attention, and the localization of mental operations in the brain, as put forward by Posner, Petersen, Fox, and Raichle (1988) are important components in this foundation. Taken together, they point to the functional components in task execution as taking place at particular local brain centers with the coordinated operation of several such centers being required to accomplish any given cognitive task. The functional task breakdown among centers is perhaps best understood for visual and auditory processing. The form that the coordination might take is of particular importance in developing a model of behaviors.

In this framework, there is the potential for a significant amount of parallel computation and it is at the same time evident that there are bounds to that parallelism. Recent work in several disciplines has suggested the forms that this parallelism might take. Edelman (1987) speaks of the degeneracy of reentrant nets in which the same functionality might be provided by several different brain centers. In the instance theory of automaticity, Logan (1988a) proposes the activation of multiple memory traces as the basis for learned responses to familiar situations. In the early stages of this process, problem solving may go on concurrently with memory retrieval. And Dennett (1991) discusses a Multiple Draft theory of consciousness in which there are always several contending syntheses of perceived events in process.

3.1. Multiple Task Management

One of the principle areas of research in the development of D-OMAR has been in the area of human multitask behaviors. In developing D-OMAR, we have sought to provide a computational framework in which to assemble functional capabilities that operate in parallel, subject to appropriate constraints, and that exhibit the multiple task behaviors of human operators. The desired behaviors have a combination of proactive and reactive components. That is, the operators have an agenda that they are pursuing, but must also respond to events as they occur. The bounds on what can be accomplished concurrently take several forms. A typical behavior may be to set aside an in-person conversation in order to respond to a telephone call, while at a simpler level, two tasks that require the use of the dominant hand can not both have access to that hand.

The core of a D-OMAR model is a network of procedures whose signal-driven activation varies in response to events that are proactively channeled to achieve the operator's goals. Currently activated goals represent the operator's proactive agenda for managing his or her tasks. The goals typically activate a series of sub-goals and procedures. The goals and sub-goals express what is to be done; the procedures are the implementation of the actions to achieve the goals. The operator's current agenda is carried out by the active procedures. However, most procedures are in a wait-state—they represent the operator's ability to cope with a changing world. They are the long-term memory about how to carry out the actions necessary to manage and respond to the changing series of events that an expert operator is likely to be confronted with.

The modeling framework provided by D-OMAR with respect to portraying multiple task behaviors differs from other human performance modeling frameworks in several important respects. Unlike EPIC (Meyer & Kieras, 1997) and SOAR (Laird, Newell, & Rosenbloom, 1987) it is not rule-based. There is not a fixed time-step with the decision process for all in–process or pending tasks revisited by processing rule sets at each step. Like MIDAS (Corker & Smith 1993), the goals and procedures that combine to represent the execution of a task are explicitly represented. However, MIDAS is also tick-based and employs an explicit decision process in the form of a scheduler invoked at each tick of the simulator clock. EPIC, SOAR and MIDAS each employ a meta-level process that reasons over executing procedures determining which is to control the agent's action in the next time step. In the models developed in D-OMAR special attention has been paid to the execution of concurrent processes and to how task contention, when it arises, might be arbitrated in the absence of a centrally located 'homunculus" or executive function.

3.2. Memory and Expectation

Procedural memory is given a central focus within the D-OMAR framework. Plans and action sequences are represented as goals comprised of sub-goals and procedures. This network of goals, sub-goals, and procedures represent what the operator's knows how to do in the world. The canonical form of the goals and procedures are important elements of the operator's long-term memory. Working memory is represented as the currently activated goals and procedures and their slot values established during their execution. The proactive disposition of the operator is set up by the goals of the operator.

In many cases, the proactive disposition of the operator improves the operator's ability to perform routinized tasks under high workload conditions involving multiple tasks competing for processing resources. However, this can also be a source of error referred to as habit capture (Reason, 1990; Norman, 1988)

Expectancy plays an important role in many of the aspects of the scenario, for example, in what the first officer writes down for the taxi sequence. Expectancy effects what you think you hear, what you decide to write down even if you heard and reported back something else. Expectations will be different for aircrews that frequently fly into the airport, those that seldom land there, and those that have never been there. The captain who is moderately familiar with the airport may be more prone to an expectation-driven failure than one who knows it "like the back of his hand" or the one who has never landed there because he/she may have erroneous expectations. We have represented the competition between what habit as episodic memory suggests is the next taxiway turn, what early intentions based on partial information is the next turn, and what is written down and read by the first officer is the right taxiway turn.

3.3. Attention

D-OMAR has the capacity to represent the reallocation of attentional resources based upon two, sometimes competing sources of control: top-down, operator-driven activation stemming from the currently activated goals and procedures and bottom-up, stimulus driven activation based on signals (typically auditory, visual, or haptic) representing sensory inputs. While most researchers agree that early stages of visual information processing occur automatically, without central capacity limitations in processing the visual field, there is ample evidence of a bottleneck at some point in the information processing stream. Attention is often assumed to select some stimuli for greater processing than others. There is still much debate in the attentional literature regarding the locus of the bottleneck (e.g., Mack & Rock, 1998; Meyer & Kieras, 1997; Yantis & Johnston, 1990) and degree of parallel processing downstream of the bottleneck (e.g., Kinchla, 1977; Treisman & Gelade, 1980; Wolfe, 1994). To date, D-OMAR has not been wed to any of these competing theories, but implementations have favored a late selection view with some degree of parallel processing even past the selection bottleneck. The key abilities to represent attentionally based information processing bottlenecks, attentional reallocation by both active goals and procedures and external stimulus inputs, combined with goals and procedure interruptions and activation of new competing procedures enables D-OMAR to predict a variety of attentionally based errors as observed in the land-and-taxi scenario.

For example, the Reason Taxonomy represents some attentional errors as slips due to inattention. In our modeling framework, the dynamic computation of priority and attention mediates contention between competing tasks. As workload increases the number of competing tasks, the probability that any given task will be deliberately or inadvertently omitted increases, amplifying the likelihood of introducing an error. The appearance of environmental stimuli such as surface lighting, signage, or other vehicles may also draw attention activating new procedures and interrupting currently active ones, often to the benefit, but occasionally to their detriment of the operator.

3.4. Discrimination

The outputs of the operator’s sensors, eyes and ears in the current model, impinge on visual or auditory processing procedures in the procedure network. Subsequent signals activate procedures that interpret the sensory inputs and lead to goal-directed responses appropriate to the evolving situation. Currently, the signals, which are interpreted by the sensory processing procedures, are clear and unambiguous, without the presence of noise. In order to effectively model the environmental conditions of the land-and-taxi scenario, the model must incorporate varying levels of stimulus discrmininability based on such factors as runway visual range, luminance levels, and distance from the stimulus. A number of studies exist in the literature that can serve as a basis for determining signal strength under the relevant environmental conditions. For example, Owens and Andre (1996) systematically examined selective visual degradation over a range of luminance levels, including civil twilight.

The variety of airport markings, signs, and surface lighting used in the surface movement guidance and control system are certainly of varying levels of discriminability. However, without the proper empirical studies on which to ground model discriminability, one might group stimuli such as signage into broad classes of discriminability and developing discrimination functions for each. In examining error as evidenced in the T-NASA experiments, descriminability did not seem to be a primary factor and hence, its consideration was deferred to a latter time.

3.5. Communication

As currently modeled, communication has a central role in the modeling of multitask behaviors. It is an important factor that can lead to several failure types. It is a source of interruption—crew activities can be interrupted and not correctly resumed, leading to failures such as missed checklist items. Alternately, an interruption may distract the captain from the observation of important taxiway signage leading to a navigation error. In particular, the untimely interruption by a ground controller in a time pressured situation lead to one class of error in the modeled taxi procedures. The interruption of the first officer’s prompt on the upcoming turn allowed a captain’s habit-based intention to be acted upon leading to an incorrect taxiway turn.

Expectancy plays a role in communication and is a likely contributor to the failure to properly act on taxi directives. They crews “know” their destination gate—it has been communicated to them by the approach controller and they can form an intent based on previous “standard” taxi sequences to the gate. This expectation and the intent that is formed are modeled explicitly and come into play when the ground controller provides the actual taxi sequence. As each turn in the taxi sequence is approached, the possible conflict between the expected and heard directive is arbitrated. When the intents are in conflict, the window for possible error is open. In setting up the trials based on the NASA event sequences, we have used a forcing function to drive the exploration of these conflicts.

4. Human Error Modeling Concepts

Human performance models have much to offer in the quest to reduce the frequency of error in tasks as complex as airport surface operations. The ability to simulate and model a “human in the loop” study will provide a significantly lower cost method for testing new procedures and equipment under a wide range of operating conditions than the typical test requiring the use of highly experienced, expensive, human operators. However, the challenge is significant. Our goal is to develop successful models of humans and equipment capabilities and their interaction within complex real-world domains that can quickly and easily be extended to new procedures and equipment. We have begun by mapping the performance space, extending the means for the expression of error in our human performance models, and then developing mechanisms that lead to humanly plausible errors based on what is known about the strengths and weaknesses of human information processing characteristics.

4.1. Error Mechanisms

Our approach to error mechanisms relies on three, largely independent, but clearly interrelated conceptions of human performance: an error taxonomy due to Reason (1990), an analysis of human information processing fallibility derived from Adams, Tenney, and Pew (1991), and the decision-making analysis of Rasmussen (1976).

We adopt the error taxonomy of Reason (1990), a simple taxonomy (see Figure 2) beginning with unintended and intended actions. Unintended actions are broken out into slips and lapses, and intended actions into mistakes and violations. This taxonomy will serve to broadly classify error types. We then seek to understand the fundamental sources of these kinds of errors, with particular attention to the kind of scenarios that involve ground-based aircraft management.

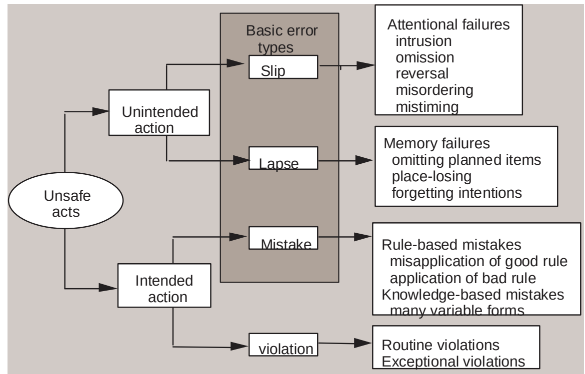


Figure 2. Error Taxonomy Derived From Reason

In our monograph (Adams, Tenney and Pew, 1991) there is an extensive discussion of the real world constraints that impact on the effectiveness of human information processing. The following quote captures the flavor of the problems:

> "We know that memory is limited. We also know that list maintenance is effortful and fallible, more so, if the list must be ordered, still more if the membership of the list must be dynamically reordered and modified during retention. These considerations suggest that proper maintenance of the queue of pending tasks would require considerable cognitive effort, even if it consisted of nothing more than a simple list of things to do. But unlike any simple list, the memory for the tasks on the queue must somehow include or point to larger complexes of knowledge and experience that underlie each. After all, it is only through access to these fuller representations that operators can update and keep track of changes in task demands, and performance parameters with unfolding events. Further it is only with reference to these fuller representations that the operator can set and adjust the logical and temporal constraints that dictate when, why and how each task can and should be carried out" (pp. 62-63).

While our monograph did not focus explicitly on error, it helps us to identify the fundamental human information processing stages most likely to be error inducing as attention, discrimination, memory, situation awareness, and planning. Attention management leads to errors resulting from a failure to attend to relevant information or from attention distraction. Attention and memory interact when a distraction away from information in immediate focus leads to forgetting of a procedural step. This was one of the contributors to the Detroit Northwest accident in 1987 (NTSB, 1988) where the step of setting the flaps for take-off was omitted due, in part, to an interruption of the Before-Takeoff-Checklist process. Attention and situation awareness interact when the aircrew or controller fail to attend to enough information or the relevant information to interpret a situation appropriately. Long term memory and planning interact when the aircrew lacks relevant knowledge about airport layout or other relevant data and fails to undertake or delays the information seeking or planning required to obtain it.

Errors rarely occur unless the crews are either too busy or not busy enough. Underload, is not likely to be a problem in aircraft ground management, but overload definitely is. Incidents rarely occur as a result of a single system failure or human error, but it is the concatenation of too many things to do combined with fallible information processing that lead to incidents and accidents. Excessive workload can result from high traffic, system failures, weather limitations, or confusing airport layouts, to name just a few sources. These factors can lead to excessive workload by generating communication requirements that take time, require additional processing capacity, and are themselves subject to fallible interpretation, or by creating complex, disproportionate demands on attention.

Operators participating in teams add another layer of complexity. Acting as team players exacerbates the multitasking demands and adds additional dimensions to the error space. Teams may work more or less well, in part, determined by their communication patterns, with good teams making effective use of implicit as well as explicit communication (Serfaty, Entin, & Johnston, 1998). Shared goals and understanding of the situation further support effective teamwork (Orasanu, 1994; Salas, Dickinson, Converse, & Tannenbaum, 1992; Serfaty et al., 1998). Failures of these team performance processes are an additional source of error to be examined.

Rasmussen (1976) (Figure 3) has laid out a decision making process idealized as activation, observation and data collection, identification of system state, interpretation of situation, evaluation of alternatives, task definition and selection of goal state, procedure selection, and procedure execution. His insight was that, of course, operators do not always follow all the steps of the process—he identified important shortcuts. The shortcuts leading to skill-based and rule-based performance, are frequently error free, but are also the source of important classes of error.

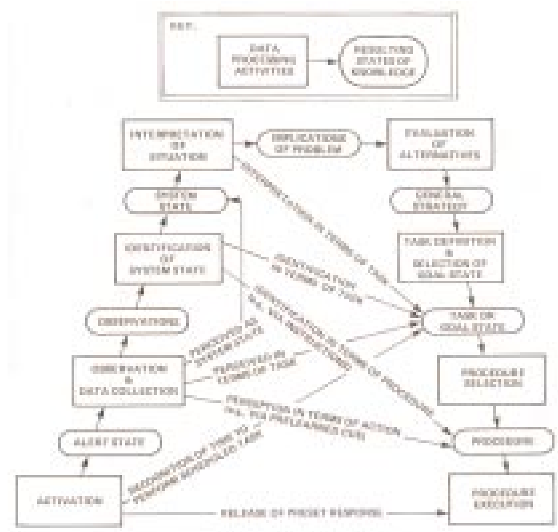


Figure 3. Rasmussen's Ladder

In 1981, we developed a conceptualization of error derived from Rasmussen's analysis that we referred to as Murphy Diagrams (Pew, Miller, & Feehrer, 1981). For a particular task context, a Murphy Diagram enumerates in a tree structure all the ways that human information processing can go wrong (and, according to Murphy's Law, if performance can go wrong, it will). Figures 4a and 4b provide two illustrations of Murphy Diagrams for developed to support this research effort. You will see that we have broken out navigation errors and failures to recognize errors in clearance (subsequently referred to simply as "clearance errors") as separate diagrams.

First, consider taxiway navigation (Figure 4a). The aircrew can either make a correct turn or turn incorrectly. If they turn incorrectly, we would argue, it can be because of a failure in spatial orientation, attention, discrimination, expectation/memory, or communication failure. In the diagram, each of these possibilities is further partitioned into the various ways in which these failure mechanisms might be manifest. For example following the "Expectation/Memory Failure" branch, we list three ways in which a memory failure could result in the wrong turn. First, the captain could have failed to write down the controller's instructions and remembers them incorrectly. Second, he/she could have written them down but made a transcription error. Third, he/she could be moderately familiar with the airport and let his expectation of where to turn to reach the gate take precedence over the controller's instruction. The other branches can be followed in a similar manner. In Figure 4b a similar diagram is presented for the case of clearance errors.

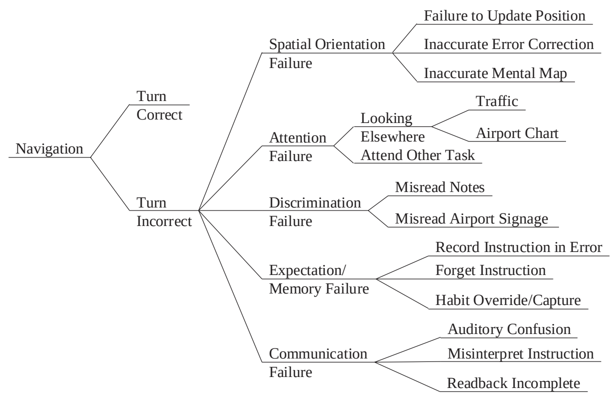


Figure 4a. Murphy Diagram for Taxiway Navigation

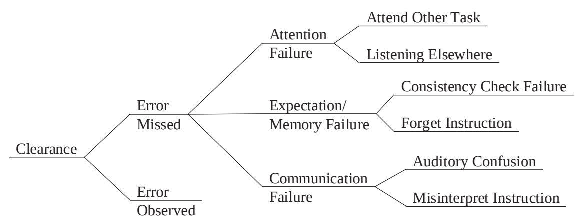


Figure 4b. Murphy Diagram for Clearance Error

It is important to point out that we do not intend to code these tree structures into our error model, rather we intend to code the mechanisms that can lead to these tree structures into the model. Thus, we create a procedure in our model where the first officer is obligated to record in writing the taxi instruction when it is received from the controller and read it back after it has been recorded. However, the model has the opportunity for competing goals to take priority and, with some probability, when the first officer is about to record the instruction, another goal might take priority. Clearly, the likelihood of this failure would be a function of the number of task demands that are competing at the time. This could result in it being recorded (incorrectly) after the readback or not recorded at all. Similarly, there is a procedure that is called when there is an opportunity to execute a taxiway turn. For the captain who has some familiarity with the airport, an expectation exists about which turn-off to take that will "compete" for priority in episodic memory with the turn-off given in the controller's instruction. These are the points at which the probabilities are set. The execution of the model then plays out to create the branching structures shown in the Murphy Diagram and produce the aircrew performance that will be used to represent the data from experimental simulation trials.

Taken together, these sources provide a framework of error that our models, at the level of the individual operators and as team players, must address.

4.2. The Relationship between Human Error and Successful Performance.

There are several critical factors essential to the model-based exploration of error in the execution of commercial airspace procedures:

- Theoretically grounded human performance models
- Models capable of the breadth of performance required by airspace procedures, including erroneous performance
- Models whose performance errors reflect human error
- An approach to locating the sequences of errors that are at the root of unsatisfactory outcomes
- An approach to finding potential remedies that successfully alter error sequence outcomes

It is our belief that the same features of human performance that lead to robust, intelligent behavior, can also induce error, in fact, behaviors that are appropriate in one context may induce errors in another. The kinds of human performance features having error implications that we are seeking to represent in D-OMAR models include:

- <u>Perception/discrimination</u>: We do not model perception at the level of primitive pattern recognition, instead our models represent perception at the level of object recognition, identification, and interpretation. Errors can occur when objects fail to be recognized because discrimination among alternative objects fails or when a sign or symbol is misinterpreted. Receipt of verbal communication is also a perceptual act. We represent each communication act as a perceptual object that can be received and interpreted correctly or incorrectly.
- <u>Memory/forgetting/interference</u>: At this point, our models are deterministic. Nevertheless, information can be lost because it exceeds memory capacity and is unable to be “chunked.” We explore forgetting by postulation specific instances of forgetting, inserting them into the scenario representation and examining the affects on subsequent activities in the scenario.
- <u>Expectancies:</u> Behavior is strongly influenced by an individual’s previous experiences. Successful performance depends on being able to anticipate what is most likely to happen next and to have formulated a plan or an “intention” for responding to it. Errors can occur when these expectancies are not realized, either because the context has changed or because the individual made a bad plan.
- <u>Situation awareness</u>: Endsley (1988) defines situation awareness as “The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future.” Situation Assessment represents the collection of all the sensory, perceptual, and cognitive activities up to the point of making a decision related to selecting and executing an action. Situation assessment leads to situation awareness.

- Intention formation: The models spawn goals as a result of some contextual event, situation or as the result of following through procedurally with a sequence of goals called for by a defined task. The "stack" of goals waiting to be initiated can be thought of as representing the intent of the crewmember. Errors arise because frequently there is competition among alternative goals or intents and they must be prioritized. Failure to act on a goal at all or at the right time, either because of excessive workload or because of failures of prioritization can easily contribute to unsafe acts.

- Execution: Crewmembers follow through with action by executing motor movements. In the aircraft, these may be movements of the yoke or actuating controls on the cockpit control panels. Errors of execution arise as the result of "slips," instances where all the decision making and intentions are correct but the wrong button is pushed or movement made. Sometimes highly automated movement sequences are completed when the context called for a less familiar one. Verbal communication is also a form of motor execution. Communication execution errors can take many forms from incomplete statements or grammatically confusing statements to failures to use the appropriate vocabulary for communicating information.

- Influences of Context: Highly practiced intellectual and physical skilled performance is context dependent. Human errors are not derived from equally likely statistically random events. Each context constrains behavior to a particular "family" of task choices and therefore induce errors specifically related to the immediate context. Each task in each context has a window of opportunity; a time when it becomes pertinent and a time when it is too late. Executing a task outside its opportunity window represents an error—sometimes the consequences are insignificant, but sometimes it leads to an unsafe act.

- Workload: When too many tasks compete for the same time window, that is, the aggregate time required exceeds time available, then crewmembers make choices or change their strategies for managing the tasks. Potentially excessive workload calls for strategic workload management (Adams, Tenney, & Pew, 1991). Whether error occurs depends on whether the task design is forgiving. Some task contexts and task designs are more forgiving than others. The ability of tasks to be cued for later action when there are too many to do at once provides a forgiving context. When they cannot be cued, the crewmember may choose to prioritize and skip less important tasks. Or he or she may choose to complete each task, but to execute it less thoroughly or completely than desired. Each of these alternatives represents an error, but not all such errors will have safety consequences. Too many tasks for the time available can also lead to "priority churn," extra time taken to reassess and reorder priorities more frequently than is really necessary.

  In addition, there are loosely coupled and tightly coupled contexts (Perrow, 1984). In tightly coupled systems task dependencies are more complex and time windows more constrained. An aircraft on final approach is in a tightly coupled context in which there is little opportunity for varying the timing or accuracy of action and each crew member must be relied on to precisely accomplish the assigned tasks. On the other hand, at cruise altitude, tasks are more loosely coupled and there is more freedom to reprogram the flight management computer, undertake less urgent or unplanned tasks, and generally review each other's actions.

- Effects of Stress: Stressors, which can include fatigue, excessive workload, high task criticality, fear of retribution for error, personal, non-job-related stress, or uncomfortable environmental conditions have the effect of amplifying potential error tendencies. Fatigue can lead to mental blinks—brief periods during which the cognitive apparatus simply closes down and information is not fully processed or acted upon. Most stressors can lead to a narrowing of focus and the opportunity to overlook potentially critical environmental events.

When viewed from these perspectives, human error can be discovered in thoughtful, detailed models of robust, successful human performance!

4.3. Putting it all Together

The component parts that make up the multi-disciplinary grounding for a human model must play together in a coherent manner that faithfully represents human multiple task performance. In this section we provide a description of how these components come together to generate the behaviors of our aircrew and air traffic controller models.

In the approach-and-landing and taxi scenarios, the captain and first officer must work together to execute the approach and ground controllers' directives, their directives must be remembered, actions must be prioritized, appropriate messages must be generated to coordinate their activities, and interrupts in the form of further directions from the approach and ground controllers must be handled. The interrupts are not unexpected, but rather meet expectations consistent with the local traffic situation. Reactive behaviors are determined within the framework of the operator's goals. In meeting their responsibilities, the captain and first officer have a significant number of cognitive tasks in process. The scenario creates a situation in which the response to demands must be carefully prioritized to achieve acceptable performance.

As a D-OMAR operator model is initiated, it is the operator's goals that are activated first. The captain is directing the approach and landing sequence and the first officer is supporting the execution of these tasks. Cognitive tasks include checking the displays for conformance or anomalies, setting the automation for final approach, confirming the approach profile, and then configuring the aircraft for landing (NASA, 2001a). Important sub-goals address identifying and maintaining an agenda of pending actions and prioritizing the execution of those actions. Additional supporting goals include communication between crewmembers to coordinate their activities, maintaining communication with the controller, and handling the transition from one controller to the next. As the plans for the operator's goals are initiated, appropriate sub-goals are invoked, and procedures are executed.

The captain and first officer's procedure networks each have a number of active nodes. Goal nodes determine the on-going proactive actions of the operator. Procedure nodes may be in a *wait-state*, typical of tasks whose function it is to anticipate and maintain vigilance for future events related to the aircraft. The activated nodes of the network form a pattern matcher with a temporal dimension such that node activation evolves in response to impinging events. The response of active nodes alters the activation of downstream nodes eventually connecting to goal-directed nodes that govern the proactive response to emerging events. The network of activated proactive and reactive nodes link functional capabilities, and in effect, generates the operator's multiple task behaviors.

As an example, an auditory input, the spoken message of a radio-based conversation, initiates an auditory procedure that collects and briefly remembers the spoken message. Early on in the auditory processing, the auditory procedure generates a signal that, in turn, initiates a non-conscious cognitive procedure that forms the propositional content of the message. And, once again, early on in the execution of cognitive (simulated) speech/language understanding process, another signal is generated that activates the several process components that, taken together, represent the operator's thoughtful management of the new conversation. The operator's management of the conversation is governed by one or more of the operator's goals. In our example, it might be the *anticipated* call from the approach controller providing the landing clearance and the preferred runway exit. Hence, the procedure network links reactive and proactive behaviors. The onset of the spoken message initiates an auditory process whose signal initiates a cascade of signals activating functional centers, which taken together, have the mix of capabilities to conduct the conversation and relate its content to the listener's purposeful goals.

The process of initiating a spoken transaction is a similar one. A thoughtful, as opposed to non-conscious, goal-related cognitive procedure generates the content of the message, a second procedure simulates the generation of the content and form that the message will take, and a third procedure represents the enunciation of the message. In the air traffic controller environment where communication is by party-line radio, the timing for initiating a transaction is important. An air traffic controller may be awaiting the response from an aircraft on a previous message as the time approaches to initiate another unrelated transaction. Barring unusual circumstances, policy dictates that the in-process transaction be completed before the new transaction is initiated. Within the model, a priority is associated with the set of procedures governing each transaction. By virtue of being active, the in-process transaction has sufficient priority to block the newly formed transaction procedure nominally of the same priority. The two transaction-governing procedures are classified as being in conflict with one another. When one task is in process and the conditions are established such that a second is a candidate to run, the conflict is resolved between the two tasks on the basis of priority. In the case outlined here it is established policy as automatically implemented by the controller that forms the basis for conflict resolution. This is representative of a conflict between thoughtful cognitive tasks that are relatively high in a goal-plan hierarchy. Conflicts can also occur lower in the complexity hierarchy. The conflict may be quite straightforward as in two procedures each requiring visual guidance in task execution and hence each needing the eyes, or each needing the dominant hand for skilled fine motor control in the execution of a task.

On-going tasks determine their own execution times and run to completion unless another procedure defined as a competing procedure with greater priority intervenes. The blocked task may be defined to resume operation at the point of interruption or at an earlier point in its execution. A thoughtful cognitive act of deciding on the next action is modeled as just that, another procedure that determines the action to follow. Importantly, a broad range of thoughtful and non-conscious decision making is represented without resorting to a central executive responsible for scheduling future actions. This approach has the potential to represent errors in the context in which they are likely to occur rather than to be probabilistically forced by the central executive.

5. Human Error Modeling Applied to Taxi Operations

NASA Ames Research Center conducted two full-mission studies of surface operations under low visibility and night conditions. The subject of the studies was the Taxiway Navigation and Situation Awareness (T-NASA) and its potential role for improving commercial aircraft airport surface operations in weather conditions to CAT IIIB while maintaining a high degree of safety (Hooey, Foyle, & Andre, 2000). The T-NASA system includes a head-up display, a head-down electronic moving map, and directional audio alerts. The studies included a series of baseline trials run without the T-NASA systems and a series of trials using various configurations of the T-NASA system. The focus of the BBN modeling effort was on the baseline trials, hence the following discussion focuses on that subset of trials.

The NASA Ames Advanced Concept Flight Simulator (ACFS) used in the studies provided a generic glass cockpit simulator with a 180-degree field of view and a high fidelity rendering of Chicago O'Hare Airport replicating the airport layout including runways, taxiways, signage, painted markings, lights, concourses, and structures (Hooey & Foyle, 2001). In the first study, 16 two-pilot commercial crews each completed six land and taxi-to-gate trials based on current operations using Jeppesen charts for navigation. Half of the trials were under low visibility conditions with runway visual range (RVR) of 700 feet, and half under night visual meteorological conditions (VMC). In the second study, 18 commercial two-pilot crews each completed three land and taxi-to-gate trials based on current operating conditions under 1000 foot RVR conditions. In evaluating these studies, Hooey and Foyle (2001) defined navigation errors as taxiing on a portion of the airport surface on which the aircraft had not been cleared and deviating from their cleared centerline by at least 50 feet. Their analysis revealed 26 navigation errors in 150 current-operation trials—errors were committed on 17.3% of the trials.

As preparation for the modeling efforts, NASA provided an extensive "Information Package for Modelers" (NASA, 2001a through 2001i). The package included materials grouped in thirteen items. Among the most useful were a series of papers related to the T-NASA experiments, a Nominal Task Sequence (NASA, 2001a) for the study scenarios, airport maps depicting routes and errors (NASA, 2001b), signage maps

(NASA, 2001h), and a description of the navigation errors (NASA, 2001i). The nominal task analysis provided a sequential listing of the tasks performed by the aircrew during the approach-and-landing and taxi flight phases. The data was derived from a sample of eight runs with median timing and range of timing data provided.

5.1. Modeling Robust Nominal Performance as a Prelude to Modeling Error

As we set out to seek the sources of error and then to model error in taxi procedures, we started by developing models that captured the robustness in aircrew procedures. The NASA Information to Modelers package included a Nominal Task Sequence (NASA, 2001a) for the T-NASA 2 baseline conditions. This was used to as the basis for the development of the approach-and-landing and taxi procedures (as outlined in Section 2) that the models of the captain, first officers, and air traffic controllers employed.

Approach-and-landing is one of the busiest phases of flight making high demands on the aircrew. But in spite of the high demands of getting the aircraft safely on the ground, it is also the time at which the first steps in the subsequent taxi operations are initiated. The crew is in the process of approaching a given runway and already know the concourse and gate toward which they will be heading. Moreover, as specified in the Nominal Task Sequence, at about eleven miles out they discuss with the air traffic controller and among themselves which runway exit they will take. As we will argue below, the crewmembers each now have in mind one and perhaps several taxi routes they might take to the gate. Once the runway exit information is in hand, the focus of attention returns to landing the aircraft and rollout.

The information provided in the Nominal Task Sequence was also used as the basis for the modeling of the remainder of the landing and rollout sequence. As the rollout sequence is completed and the aircraft approaches the designated runway exit, the taxi sequence is initiated. The first officer provides information to the captain on their position relative to the preferred exit based on notes taken when the preferred exit was agreed on. He/she then informs the controller that the aircraft is clearing the runway, both crewmembers then switch their radio frequency, and the first officer contacts the ground controller. At this point, the ground controller provides the crew with the taxi route to the gate and the first officer writes down the taxi route.

It was at this point that we encountered the first instance of a requirement for a coping strategy. Many of the high-speed exits at O'Hare have a very short run to the first intersection and taxiway routings can be unusually lengthy. We encountered this first when modeling a landing on runway 9R using high-speed exit M6 with an immediate left turn onto taxiway M. The first officer was head-down writing out the taxi directives and was late in providing information to the captain on the upcoming immediate turn. At this point, the captain was also listening to the taxi routing and could go with what he/she heard or slow the aircraft and get confirmation from the first officer on the upcoming turn. The coping strategy that we modeled had the captain go ahead with the turn as heard and notify the first-officer of the turn as it was executed.

The process for each subsequent turn in the taxi sequence followed the same pattern. As a turn was completed, the first officer would consult his/her routing notes and the Jeppesen airport diagram, and then prompt the captain on the taxiway and direction for the upcoming turn. As expected, the modeled nominal process proved very robust. By simply changing the routing that the ground controller provided, the captain and first officer were able to execute any desired taxi routing. With these robust aircrew processes in place, the challenge was to model taxi sequences that produced errors consistent with those in the baseline T-NASA experiments.

5.2. Local and Global Situation Awareness

As the captain and first officer meet their responsibilities during taxi operations, the inherent nature of the tasks that they perform provide them each with a different sense of their immediate location and their location with respect to their assigned taxi routing. They each achieve and maintain different levels of local and global situation awareness. And indeed, if they are working well as a team, they will strive to fill each other's gaps in awareness. In building the aircrew models, we felt that it was essential to reflect this level of teamwork.

In modeling the captain's tasks, he/she was modeled as predominantly head-up during taxi operations. He/she announced each turn as it was executed to keep the first officer informed of their immediate location during such periods as the first officer might have been head-down. Meanwhile, the first officer worked with a Jeppesen airport diagram and kept written notes on the runway exit and taxiway routing. The use of these aids enabled the first officer to provide the captain with a more global view of their taxiway routing than would have otherwise been available. As modeled, the first officer noted upcoming turns that had short lead times, intermediate taxiway crossings for turns with longer lead times, and made special note of the idiosyncratic point on O'Hare taxiway D where it is necessary to execute a turn to stay on taxiway D. The teamwork skills of the modeled aircrews were effective in repairing gaps in one another's situation awareness. One effect of providing this level of detail in good crew performance was of course to make the taxiway procedures just that much more robust and error that much less likely.

5.3. Making the Wrong Turn at an Intersection

The particular process that produced the errors of interest was the preparation for and execution of the next turn in the taxi sequence as governed by the captain. As modeled, the captain, if left to his/her own resources, must rely on his/her memory of the taxi sequence as conveyed by the ground controller as the aircraft cleared the landing runway. However, the captain gets significant support in this task from the first officer. The first officer takes notes on the taxi sequence as it is received from the ground controller and will, under nominal conditions, prompt the captain with the name of the next taxiway and the direction of the turn required. Our model makes the task of recording the taxi route easier than it actually is in the real world—we did not realize that the ground controller provides only the name of each taxiway and not the direction to turn at the intersection. In our model, the ground controller provided both the taxiway name and the direction of the turn.

Under nominal conditions, the first officer will prompt the captain on the upcoming turn and one can reasonable expect that the captain will correctly act in accordance with the prompt. Acting counter to the prompt is an error possibility that we did not pursue. Hence, to find a source for making a turn error at an intersection we had to construct reasonable scenarios in which the first officer was otherwise occupied and unable to provide the prompt in a timely manner and of course identify an underlying reason for a mistake on the part of the captain. The events that prevented the first officer from providing the prompts are discussed below in the sections providing details on the error scenarios. Here, we examine possible sources for the errors committed by the captain in executing the incorrect turns.

5.4. Intention-to-Act

A classical view of the taxiway process might be that in approaching a turn the captain has a planning problem whose resolution is then followed by plan execution. Do we in fact need to make a turn at the upcoming intersection and if so, which way? There might be a schema in place for executing the next turn with slots to be filled in for the name of the next taxiway and the direction to turn. In this view of the process, error might come about by incorrectly filling the slot for the next taxiway name, but more probably, the slot for the direction of the turn to make.

We would like to argue in favor of an alternate view in which there are typically several *intentions-to-act* concurrently in process. The intentions may be established at different points in time. One or more of them may lead to a correct turn or to making an error at the intersection. A winner–take-all process leads to the execution of one of the intentions-to-act and the correctness of the outcome is the product off the winning intention. At the point of execution, the remaining intentions cease to contend. We label the process intention-to-act to suggest that the process is not the product of a conscious decision process—it is not a deliberative planning process followed by a plan execution process. There is the immediacy of an automatic, atomic process rather than a sequential process of planning and acting. Each of the intentions-to-act is instantiated with established slot values, rather than with unfilled slot values to be filled by a deliberative process.

Most of the time there is more than one intention-to-act with the contention resolved in a winner-take-all competition. In the nominal case where the first officer has provided the correct prompt for the turn, the turn is, most likely, simply executed in response to the prompt and most likely in accordance with a pre-existing intention. In lieu of the prompt from the first officer, the captain will act on a pre-existing intention that might lead to the execution of his/her intention to turn or alternatively to pause and query the first officer on the next turn. (We have not pursued the case of the captain’s slowing or stopping the aircraft and querying the first officer.) That is, most of the time in the taxi environment it is reasonable to expect that the captain has an intention-to-act in place and ready to be acted on.

Rather than having a single planning process with slots to be filled from various sources that is followed by a plan execution step, there are multiple intentions-to-act with selection through a non-conscious winner-take-all process. Each of the intentions-to-act has a complete set of immediately filled slots. In the following section, we provide the reasoning supporting this viewpoint.

5.4.1 Multiple Intentions-to-Act

At this point, we want to build the case for the idea that in performing relatively simple tasks like correctly executing the next taxiway turn, there may be several competing intentions-to-act. Most may be automatic processes that require little or no conscious deliberative thought. They may emerge from different ongoing processes competing in a winner-take-process to determine the action taken. Occasionally, the winner will determine an action that is in error. During the course of this study, we have attempted to identify some of the sources for these intentions and to provide reasoned explanations on why the errors emerge.

For most of us there are a broad range of everyday activities that we perform quickly and effortlessly—they appear to be automatic and involve little thought or conscious awareness (Logan, 1988a; James, 1890). Logan (1988a) characterizes this automaticity, the execution of these activities, as fast, effortless, autonomous, stereotypic, and unavailable to conscious awareness. That is, we experience them as fast, effortless, stereotypic, and unavailable to conscious awareness. They are autonomous in the sense that the acquisition of these skills comes about independent of any deliberate intention to learn them.

Logan (1988a) developed the “Instance Theory of Automaticity,” a theory for how automatization is constructed. The theory was developed in part through a series of experiments in learning alphabet arithmetic, learning to solve problems of the type “A+2=?” where the answer is “C.” Initially most people solve these problems by explicitly counting out the required steps through the alphabet—they employ an algorithm which they step through. With experience they “learn or remember” some of the answers and for some of the problems they have addressed before. There is then a fast and ready answer—the explicit problem solving process is not necessary this time.

Logan suggests that *each* learned instance is remembered. When presented with a new problem, there is a concurrent attempt to access a remembered instance of a previous solution *and* an explicit problem solving computation. The memory access is a comparatively fast process, the deliberative process comparatively slow. If the memory access is successful in retrieving a solution, there will be a rapid response to the posed problem. If the memory access is not successful, the response will be slower. Through experience, more and more solutions are acquired and at some point, the deliberative process is simply not a contender in the winner-take-all process. For any given problem, there may be several remembered solutions. Due to the remembering of each solution instance, there may potentially be several correct retrievals. It is the one that is first retrieval that determines the time required to solve the problem.

Logan (1988b) further argues that the memory traces that support automaticity may well support declarative as well as procedural knowledge. Logan (1988b) suggests that we “look more broadly for automatic processes. They need not be restricted to procedural knowledge or perceptual-motor skill but may permeate the most intellectual activities in the application environment." Here we are suggesting that the captain's procedures for addressing the next turn in the taxiway sequence may sometimes be characterized as automatic and that while these will often lead to correct behaviors, they may sometimes lead to errors such as those seen in the T-NASA experiments.

5.4.2 Intentions-to-Act as a Source of Error

Our review of the NASA-provided data on the T-NASA experiments pointed to two important factors that we felt deserved particular attention in our modeling effort. NASA (2001i) identified the importance of the location of the destination gate and its relation to the taxi route. Five errors occurred in 48 instances of required turns *away from* the shortest route to the concourse gate while only seven errors occurred in 534 instances of turns *toward* the concourse gate. At any given intersection, the aircrews had a bias to turn toward their destination concourse gate. When the correct turn was one away from the concourse gate, there was a greater tendency toward making an error. The second observation was the straightforward one that time pressure can lead to error. There was a greater chance of error when a second turn in the taxi sequence closely followed the previous turn. The time pressure of a second turn closely following a first turn was an important factor in each of the errors that we generated in the modeling effort.

To date, four sources of intention-to-act have been identified and modeled. Each was initiated at a different point in the taxi sequence. The first is episodic memory—in part a source of autobiographical memory and a source for habit-based actions. Similar situations have been encountered in the past and we have a ready source of responses that have worked just fine. These are responses that in the past have proven successful and are generally able to carry us through most of the activities of the day. When they fail this is what Reason (1990) refers to this as "strong-but-wrong." In our particular case, the aircrews have a history of previous landings at Chicago O'Hare.

A second source of intention-to-act is context-based expectation, driven by partial knowledge. Explicit partial information provided in the current situation prompts a particular intention. Within the taxi-framework, the captain knows the location of the concourse gate and based on this knowledge may reasonably have an expectation that the next turn will take them on the shortest route to the gate. These particular situation-specific information points are sufficient to set up an intention for the next turn.

The third source of intention-to-act is the remembrance of the taxi sequence provided by the ground controller as the aircraft exited the landing runway.

The fourth source of intention-to-act in the taxi-framework and the best-grounded source of intention is the explicit prompt by the first officer based on written notes on the taxi directives from the ground controller. In the nominal case, the first officer's prompt will match the captain's intention and will lead to error free performance.

We model the contention between these intentions as a winner-take-all process mediated by priority and explored the impact of varying the priorities of the contending intentions. Within the winner-take-all framework, at the winning intention's transition from intention to action, the remaining intentions cease to contend—within the framework of the model, their procedures fail. Timing of the events that drive the intentions determine how they play out producing successful behaviors or mistakes in behavior that can lead to an incorrect turn on the taxiway. In particular, to provide a window for error to occur, it was necessary to set up realistic event chains that prevented the first officer from providing the prompt to the captain on the next turn.

As we have suggested, the team-based nature of the taxi procedures makes them very robust and the challenge has been to create situations in which mistakes will lead to error. This effort focused on two error sequences that each required two turns in rapid sequence, in one case, immediately after a high-speed exit from the runway, and in the second case, later in the taxi sequence. For case one, there were two instances of the same error as crews took high-speed exit M7 from runway 9 right. At the first intersection after the high speed exit each captain turned left toward the concourse gate rather right as directed by the ground controller. In the second case, there were two scenarios that shared a similar turn sequence: after turning onto taxiway F in the first instance and M2 in the second instance, there was a quick right turn onto taxiway B. In each of the scenarios, one of the captains turned left rather than right. The errors were noteworthy because the captains each turned *away* from their intended concourse gate rather than toward the gate as directed.

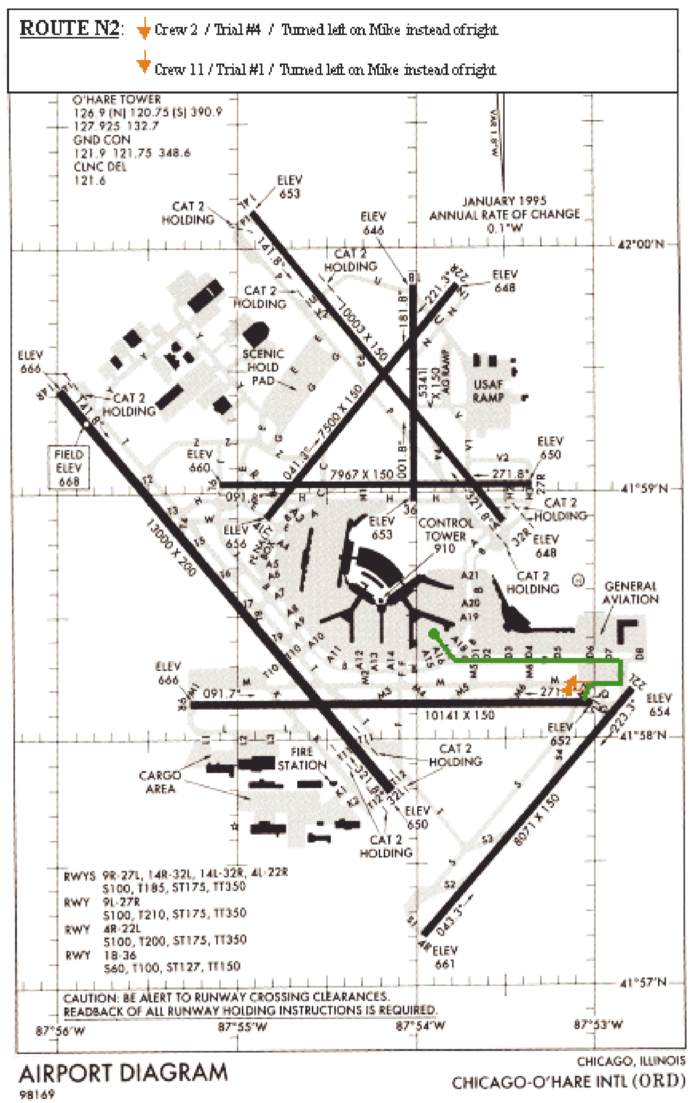


Figure 5. Errors when Turning Toward the Gate (NASA 2001b)

5.4.2.1 *Error Driven by Expectation Based on Partial Knowledge*

Our hypothesis is that the incorrect turn following the high-speed exit (see Figure 5) was driven by the captain's expectation that the shortest route to the gate was the route to be taken. (The small arrows that denote the errors in Figure 5 indicate the incorrect left turns taken just after the high-speed runway exit. They are in red when viewed in color—in grayscale, they may be difficult to make out.) The intention-to-act arouse at the point of the early discussion of the runway exit with the approach controller and the first officer. At this point, the captain knew the runway exit and the concourse gate, and might reasonably have expected to turn left from the high-speed exit at taxiway M taking him/her toward the gate. It became one intention contenting to be executed at the first turn after exiting the active runway.

As the scenario played out in the nominal case, the first officer completed the task of taking notes on the taxiway sequence and then prompted the captain on the first turn following the runway exit. The first officer's prompt triggered a new, contending intention-to-act on the captain's part. The new intention may or may not have been consistent with preexisting contentions. In the nominal case, it dominated and the captain turned right correctly. Given the correct prompt by the first officer, we deemed it highly unlikely that the captain would incorrectly execute the turn.

To open a window for an error to occur, it was necessary to construct a situation that reasonably occupied the first officer preventing him/her from providing the captain with the explicit prompt on the upcoming turn. The very short run to the first turn after the high-speed exit was the essential factor. The first officer was already busy taking notes on the taxiway routing. Indeed, in some scenarios the taxiway routing was so lengthy that in the nominal case the first officer was still taking notes as the first turn was executed. In this scenario, this was not the case, hence a "mistake" was needed to additionally task the first officer. The failure to preset the radio frequency for the transfer to the ground controller provided the delay. The few seconds necessary to set the new radio frequency provided enough delay to prevent the first officer from prompting the captain on the turn. This was a mistake on the part of the aircrew in the sense that it is always incumbent upon them to complete an action at the earliest available time, rather than risk a situation such as this in which there are contending tasks in process.

Let us recap the captain's intentions-to-act as the aircraft approached the first turn onto taxiway M after the high-speed exit on taxiway M7 following the landing on runway 9R. The first officer has been otherwise occupied and has not provided the captain with the explicit prompt on the upcoming turn. Based on the coping strategy described earlier, the captain might have a correct intention-to-act based on having attended to the ground controller's taxi directive and an incorrect intention based on the expectation of receiving a shortest route to the concourse gate. Much of the time the coping strategy might be expected to win the winner-take-all competition and lead to a correct turn—some of the time the expectation-based intention-to-act might be acted upon leading to a taxiway error. Hence, a reasonable, grounded source for an error consistent with the T-NASA experiments has been identified and modeled.

5.4.2.2 *Error Driven by Habit*

The second scenario examined the surprising cases in which an aircrew incorrectly turned away from the shortest course to the gate (see Figure 6). (In Figure 6 the small arrows denoting the errors indicate the incorrect left turns taken just after the short north-bound segments near the center of the airport diagrams. They are in red when viewed in color—in grayscale, they may be difficult to make out.) The basic intention

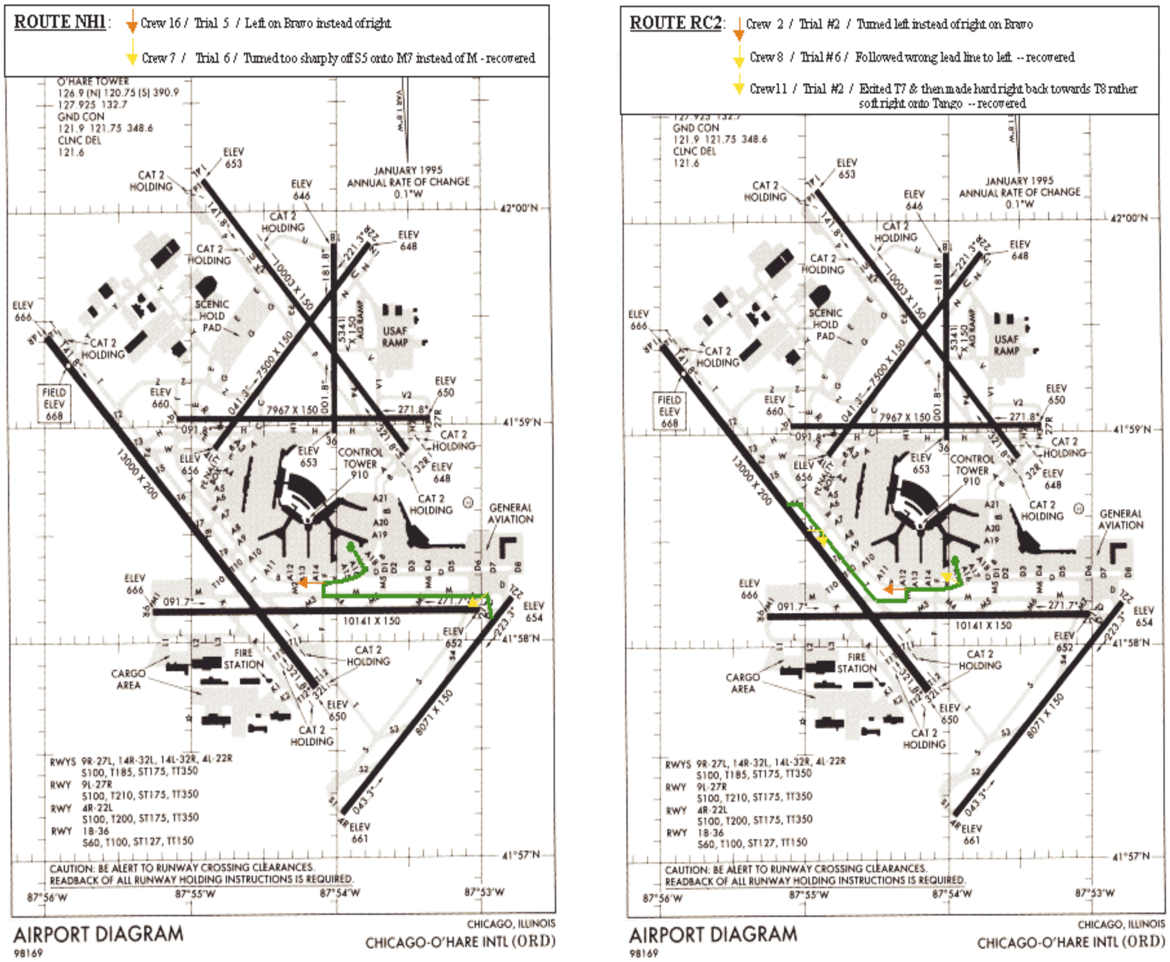


Figure 6. Errors when Turning Away from the Gate (NASA, 2001b)

to take the shortest route to the gate would have led to the correct behavior, yet it was not the acted upon intention. There were two instances of this error at similar intersections. In the first case, the aircraft was proceeding north on taxiway F and had been instructed to turn right onto taxiway B, but the captain turned left instead. In the second case, the aircraft was proceeding north on taxiway M2 and had been instructed to turn right onto taxiway B, but the captain turned left instead. We speculated that a crew whose company gates were on the opposite side of the airport from those required by the scenario might incorrectly turn toward their company gates exhibiting an error based on long established habit. Requiring an aircrew to proceed to a gate opposite in direction from their company gates might be considered an artifact of the particular scenario, but in a commercial air travel environment that has seen many company mergers and failures it is not uncommon for aircrews to find themselves working for new companies with new gate locations.

The turn at which the errors occurred closely followed a previous turn creating a time-pressured situation. Once again, we manipulated the situation such that the first officer was not able to provide a timely prompt to the captain on the upcoming turn. Conflicting taxiway traffic was present on the first officer's side of the aircraft during the approach to the first turn. The first officer informed the captain of the presence of the traffic and continued to monitor the other aircraft. Consequently, the first officer was delayed in going head-down to review his/her notes on the upcoming taxiway turn and checking the airport diagram. Following the delay, the first officer's prompt on the upcoming turn was immediately interrupted by a message from the ground controller directing the other aircraft to hold short of the upcoming intersection allowing the first aircraft to proceed with the turn. Very slight changes in timing of the interruption would have opened the window for a successful timely prompt.

In the absence of the prompt, there were still multiple intentions-to-act. As modeled there were intentions-to-act based on the remembrance of the ground controller's taxi directive and on habit based in episodic memory. When the captain's habit-based intention-to-act won the winner-take-all competition and was acted upon, the error was committed. An informal post hoc analysis of the human subject trial error provided support for the speculation that the model represented (B. Hooey, personal communication).

5.5. Heuristically Guided Search of the Error Space

The incidence of error in the current-equipment T-NASA experiments was strikingly high when compared to the typical behaviors of professional aircrews. In general, the low frequency of mistakes and the even lower frequency of mistakes combining to produce errors renders a simple stochastic exploration of the behaviors space impractical. The robustness of aircrew team procedures that employ checking and cross-checking of critical actions means that most mistakes will be caught further compounding the search task. Estimating error frequency for error types can also be a problem. For some errors (e.g., discrimination of taxiway signage), their frequency might be reasonably estimated, for others (e.g., the onset of a particular intention-to-act) it is more difficult.

Timing is also critical. Very small variations in timing can open or close out the window in which an error might occur. Timing was particularly critical in the scenario in which the habit-based error occurred. The combination of the demand on the part of the first officer for head-up time to monitor the approaching traffic and the precise moment of the ground controller's interruption of the first officer's prompt for the upcoming turn was necessary to open the window to error. It might well have been possible to generate many hundreds of runs slightly varying several of the timings and never have produced a single habit-based error.

To address this problem, we have employed a heuristically guided search of the space in which forced sequences of mistakes are generated looking for those that lead to error. The errors produced to date are initial examples of the product of such a process. We have identified several novel, potential sources of mistakes and worked to create situations in which they might reasonably be expected to occur. We have taken advantage of the time pressure inherent in the closely spaced turn sequences to manipulate the timing of events to construct sequences of mistakes that do in fact lead to error. For the present, this heuristically guided exploration of the error space has been manipulated by hand. In the future, we would like to move toward a more automated exploration of the error space.

6. References

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1991*). "Strategic workload and the cognitive management of advanced multi-task systems."* Wright Patterson AFB, OH, Crew Systems Ergonomics Information Analysis Center.

Brachman, R. J., & Schmolze, J. G. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive Science, 9,* 171-216.

Corker, K. M., & Smith, B. R. (1993). An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation automation. In *Proceedings of the AIAA Computing in Aerospace 9 Conference,* San Diego, CA.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little, Brown and Company.

Deutsch, S. E. (1998). Interdisciplinary foundations for multiple-task human performance modeling in OMAR. *Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society,* Madison, WI.

Deutsch, S. E., & Adams, M. J. (1995). The operator model architecture and its psychological framework. *Proceedings of the 6th IFAC Symposium on Man-Machine Systems*. Cambridge, MA.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (Volume 1, pp. 97-101). Santa Monica, CA: Human Factors Society.

Foley, D. C., Andre, A. D., McCann, R. S., Wenzel, E., Begault, D., & Battiste, V. (1996). Taxiway navigation and situation awareness (T-NASA) system: Problem, design philosophy, and description of an integrated display suite for low visibility airport surface operations. *SAE Transactions: Journal of Aerospace, 105,* 1411-1418.

Hooey, B., Foyle, D., & Andre, A. (2000). Integration of cockpit displays for surface operations: The final stage of a human-centered design approach. In *Proceedings of the AIAA/SAE World Aviation Congress*. Warrendale, PA: SAE International.

James, W. (1890). *Principles of Psychology*. New York: Holt.

Kinchla, R. A. (1977). The role of structural redundancy in the perception of visual targets. *Perception & Psychophysics*, *22*, 19-30.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence, 33,* 1-64.

Logan, G. D. (1988a). Toward an instance theory of automatization. *Psychological Review, 95,* 492-527.

Logan, G. D. (1988b). Automaticity, resources, and memory: Theoretical controversies and practical implications. *Human Factors, 30,* 583-598.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Meyer, D. E., & Kieras. D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104,* 3-65.

NASA. (2001a). Nominal task sequence.

NASA. (2001b). Routes and errors: Locations and types.

NASA. (2001c). Visual scene information.

NASA. (2001d). Individual pilot differences spreadsheet.

NASA. (2001e). Communication data.

NASA. (2001f). Pilot workload and situation awareness data.
NASA. (2001g). T-NASA2 navigation error classification.
NASA. (2001h). Chicago O'Hare Airport signage maps.
NASA. (2001i). Enhanced descriptions of off-route navigation errors.
Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action*. London: Lawrence Erlbaum.
Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.
NTSB (1988) *Aircraft Accident Report – Northwest Airlines, McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne Co. Airport, Romulus, MI, August 16, 1987,* Washington D.C.: NTSB.
Orasanu, J. M. (1994). Shared problem models and flight crew performance. In N. McDonald, N. Johnston, & R. Fuller (Eds.), *Aviation psychology in practice*. Aldershot, UK: Ashgate.
Owens, D. A., & Andre, J. T. (1996). Selective visual degradation and the older driver. *International Association of Traffic and Safety Sciences, 20,* 57-66.
Perrow, C. (1984) *Normal Accidents*, New York: Basic Books.
Pew, R. W., Miller, D. E., & Feehrer, C. E., (1981) "Evaluation of proposed control room improvements through analysis of critical operator decisions." *Electric Power Research Institute Report NP-1982*, Palo Alto, CA
Posner, M. I., Peterson, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science, 240,* 1627-1631.
Rasmussen, J. (1976). Outlines of a hybrid model of the process plant operator. In T. B. Sheridan & G. Johannsen (Eds.), *Monitoring behavior and supervisory control*. New York: Plenum.
Reason, J. (1990). *Human Error*. Cambridge: Cambridge University Press.
Salas, E., Dickinson, T. L., Converse, S. A. & Tannenbaum, S. I. (1992). Toward and Understanding of Team Performance and Training. In Robert W. Swezey and Eduardo Salas (Eds.), *Teams: Their training and performance*. Ablex Publishing Company, Norwood, NJ.
Serfaty, D., Entin, E. E., & Johnston, J. H. (1998). Team coordination training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 221-245). Washington, D. C.: American Psychological Association.
Treisman, A. M. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97-136.
Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202-238.
Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 135-149.